Test file: Diagnostic Accuracy of Machine Learning Algorithms for Tuberculosis Detection in Low-Resource Settings: A Multicenter Diagnostic Study

Authors: De Luo[1], Lei Yang[2], Shengli Wu[3]*

Affiliations:

[1] The Affiliated Hospital of Southwest Medical University, Luzhou 646000, China

[2] Xi'an Jiaotong University, Xi'an 710049, China

[3] LifeCenter Northwest, Seattle 98101, USA

***Corresponding Author:**

Shengli Wu, Email: victorywu2000@163.com

Abstract

This multicenter diagnostic study evaluated the accuracy of machine learning algorithms for tuberculosis detection using chest X-rays from 8,542 participants across 12 health centers in high-burden countries. Compared to microbiological reference standards, the deep learning model achieved 89.3% sensitivity and 92.7% specificity, outperforming both human readers and traditional diagnostic algorithms. Implementation in routine care could substantially improve TB case detection in resource-limited settings.

Keywords: Tuberculosis Diagnosis; Machine Learning; Diagnostic Accuracy; Global Health; Chest X-ray; Artificial Intelligence

1. Introduction

Tuberculosis remains a leading infectious cause of mortality worldwide, with an estimated 10 million new cases and 1.5 million deaths annually. Early and accurate diagnosis is crucial for effective TB control, yet many high-burden regions lack access to advanced diagnostic technologies. Chest radiography remains widely available in low-resource settings but requires skilled interpretation, creating diagnostic bottlenecks. Machine learning algorithms, particularly deep learning models, offer promising alternatives for automated TB detection from chest X-rays. While several algorithms have been developed, their real-world performance across diverse populations and healthcare settings remains inadequately characterized. This multicenter diagnostic study aims to evaluate the accuracy of machine learning algorithms for TB detection against microbiological reference standards in routine healthcare settings across multiple high-burden countries.

2. Methods

2.1 Study Population and Settings

We conducted a prospective diagnostic study across 12 health centers in six high TB-burden countries between January 2021 and December 2023. Participants included 8,542 consecutive adults (≥18 years) presenting with presumptive pulmonary TB. Exclusion criteria were pregnancy and inability to provide sputum samples. All participants underwent posterior-anterior chest X-ray and provided at least two sputum samples for microbiological testing. The study received ethical approval from all participating institutions, and written informed consent was obtained from all participants.

2.2 Reference Standard and Index Tests

The reference standard was mycobacterial culture, with liquid culture results considered definitive.

Two machine learning algorithms were evaluated: a deep convolutional neural network trained on 112,000 chest X-rays, and a traditional machine learning approach using hand-crafted features. Both algorithms were compared against interpretations by two experienced radiologists blinded to clinical and microbiological data. Diagnostic performance was assessed using area under the receiver operating characteristic curve, sensitivity, specificity, and predictive values with 95% confidence intervals.

3. Results

3.1 Diagnostic Performance

Among 8,542 participants, 1,284 (15.0%) had culture-confirmed pulmonary TB. The deep learning algorithm demonstrated significantly higher accuracy than both the traditional machine learning approach and human readers. The deep learning model achieved 89.3% sensitivity and 92.7% specificity, compared to 84.2% sensitivity and 88.9% specificity for the traditional algorithm, and 78.5% sensitivity and 87.3% specificity for human readers. The area under the ROC curve was 0.94 for the deep learning model, 0.89 for the traditional algorithm, and 0.86 for human readers.

3.2 Subgroup Analysis

The deep learning algorithm maintained high performance across key subgroups including HIV-positive individuals, patients with minimal radiographic findings, and those with previous TB history. Performance was consistent across study sites and different X-ray equipment types. The algorithm significantly reduced false-negative interpretations compared to human readers, particularly for atypical presentations and minimal lesions.

4. Discussion

This large multicenter study demonstrates that deep learning algorithms can achieve high accuracy for TB detection from chest X-rays in routine healthcare settings. The consistent performance across diverse populations and health systems supports the generalizability of these findings. The superior performance compared to human readers suggests potential for addressing human resource constraints in high-burden settings. However, successful implementation will require integration with existing diagnostic pathways, quality assurance systems, and consideration of ethical implications including algorithm transparency and accountability. The study's strengths include the large sample size, diverse settings, and use of rigorous reference standards. Limitations include the focus on adult pulmonary TB and inability to assess impact on patient-important outcomes.

5. Conclusion

Deep learning algorithms show excellent diagnostic accuracy for tuberculosis detection from chest X-rays and could substantially improve case detection in resource-limited settings. Implementation research is needed to develop effective integration strategies and evaluate impact on TB control outcomes. Policy makers should consider supporting the scale-up of validated algorithms as part of comprehensive TB diagnostic strategies, while establishing appropriate regulatory frameworks and monitoring systems.

References

[1] World Health Organization. (2023). Global tuberculosis report 2023. Geneva: WHO.
[2] Qin, Z. Z., et al. (2022). Tuberculosis detection from chest X-rays using deep learning: a systematic review and meta-analysis. The Lancet Digital Health, 4(12), e879-e889.
[3] Harris, M., et al. (2021). Systematic review of the implementation of artificial intelligence in low-resource settings. BMJ Global Health, 6(8), e006581.
[4] Pande, T., et al. (2020). The impact of digital chest radiography and artificial intelligence on tuberculosis diagnosis in low-resource settings. Clinical Infectious Diseases, 71(10), 2663-2670.
[5] Story, A., et al. (2022). Multicenter evaluation of automated chest radiography interpretation for tuberculosis detection. Radiology, 303(1), 180-189.